

Delovni sklop 2: Govorne tehnologije

Darinka Verdonik, Marko Bajec, Simon Dobrišek

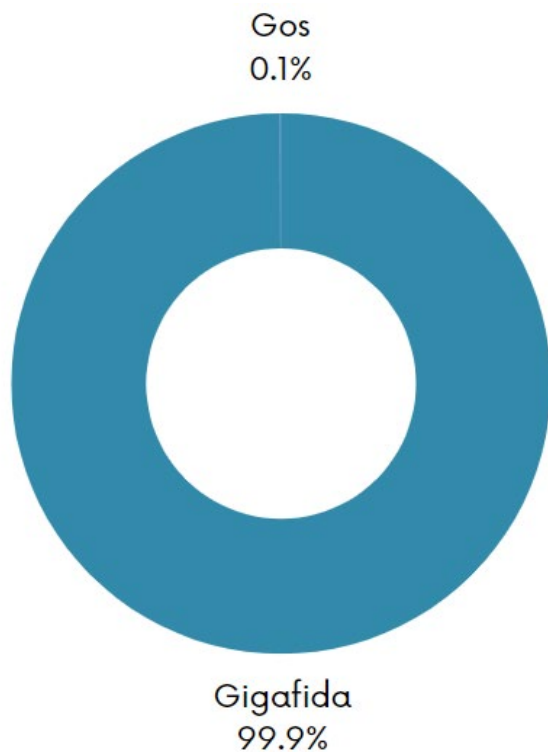
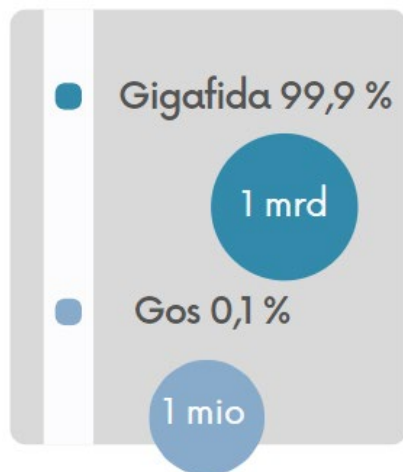
www.eu-skladi.si



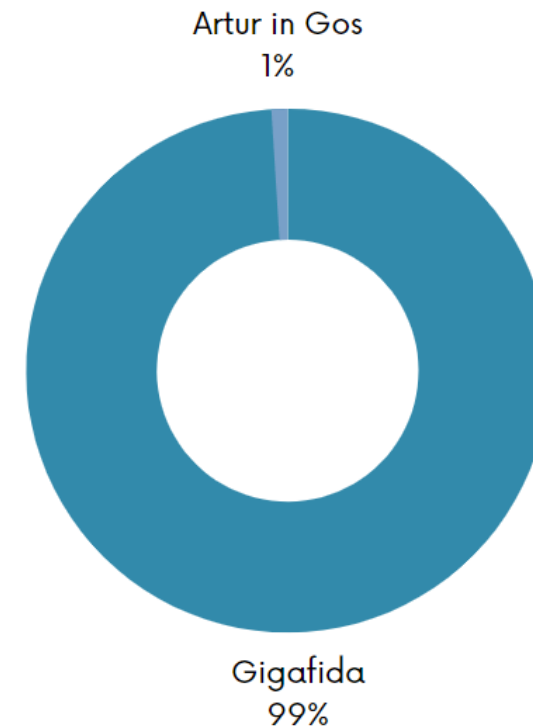
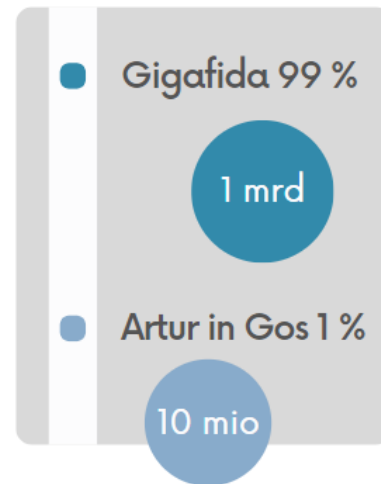
DS2: Govorna baza ARTUR



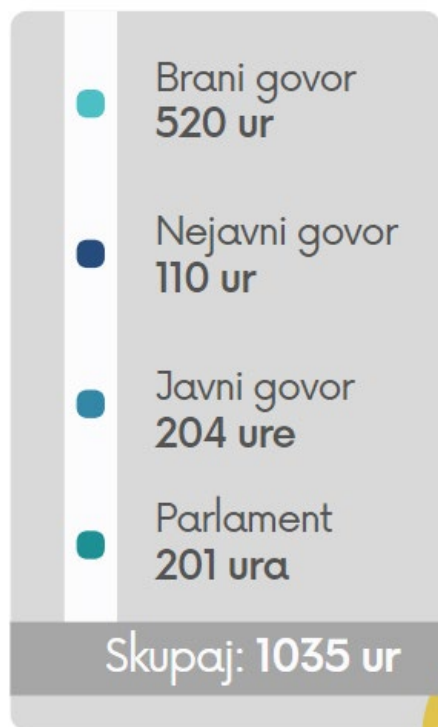
Stanje pred RSDO
2019



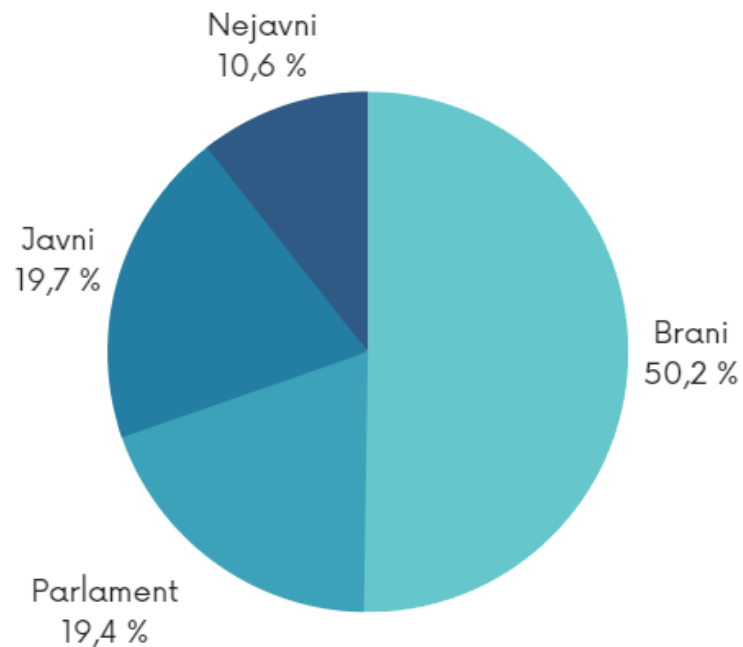
Stanje po RSDO
2022



DS2: Govorna baza ARTUR



Sinteza
50 ur



Transkripcije
840 ur

ARTUR =
Avtomatsko
Razpoznavanje
Razvoj
slovenščine v
digiTalnem
okoljU

DS2: Govorna baza ARTUR

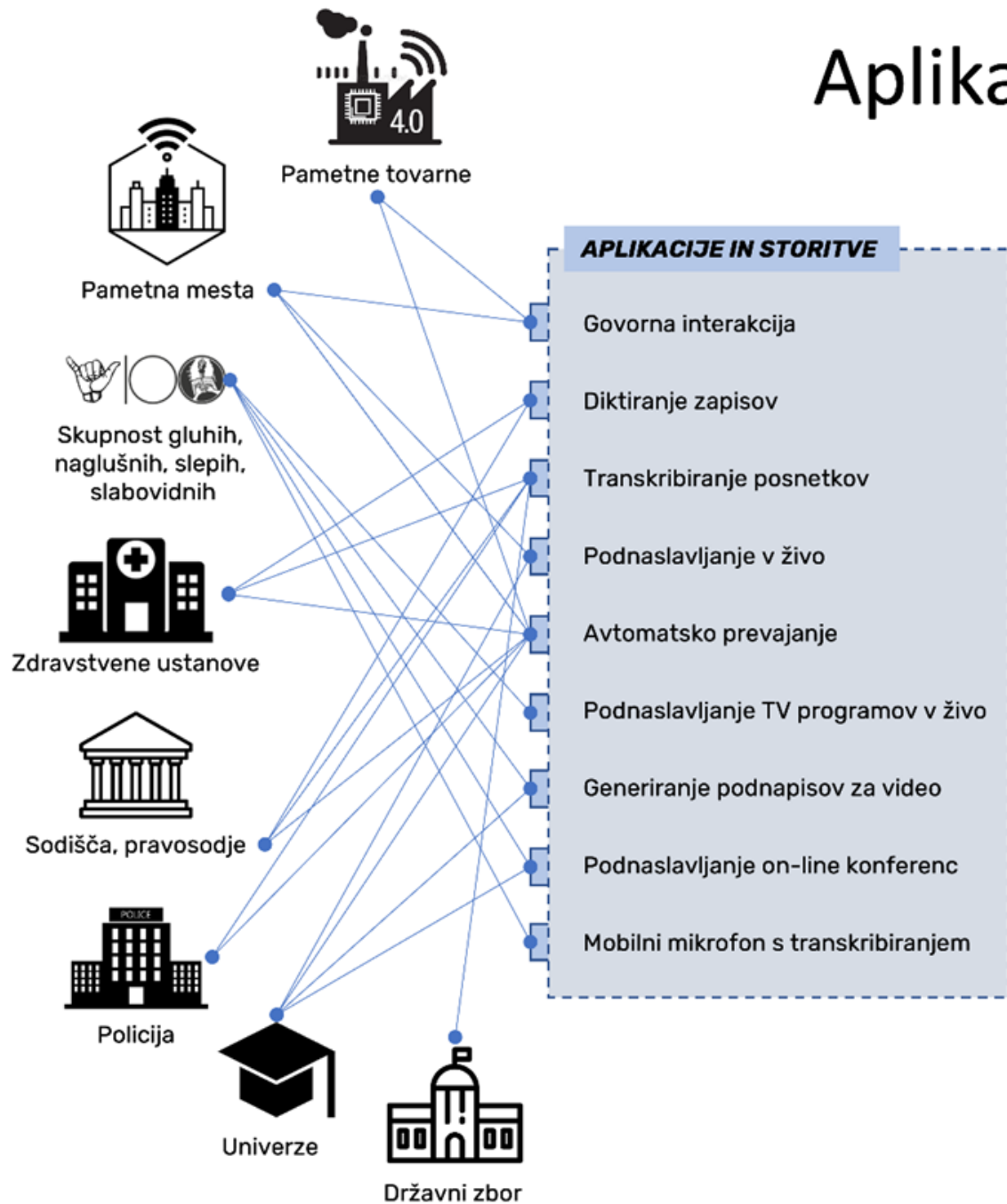


Sodelavci	Gradivodajalci
Univerza v Mariboru - FERl	STA
Univerza v Ljubljani - FE in FRI	Univerza v Ljubljani
ZRC SAZU	Univerza v Mariboru
STA	ZRC SAZU
Alpineon	Radio Štajerski val
Študenti Univerze v Ljubljani in Univerze v Mariboru	Arnes
3 zunanji izvajalci	Terenski posnetki

25.000-30.000 delovnih ur,
več kot 250 oseb

1531 govorcev iz 157 občin

Aplikacije – tehnologije – viri



GOVORNE TEHNOLOGIJE

OSNOVNE

RSDO Avt. razp. za posnetke govora

Avt. razp. govora v živo

RSDO Avt. prevajalnik iz slov. v tuje j.

Sintetizator govora

PODPORNE

RSDO Avtom. postavljanje ločil

RSDO Avtom. postavljanje velikih zač.

RSDO Fonemizator

Avtomatsko postavljanje naglasov

RSDO Sintaktični normalizator

RSDO Vektorske vložitve besed/fraz

RSDO Lematizacija, tokenizacija, oblikoslovno označevanje...

GOVORNI VIRI

Govorna baza – posnetki govora, branega in spontanega za akustično modeliranje **RSDO**

Transkripti govorne baze ortografski in standardizirani **RSDO**

Studijski posnetki profes. govornika za potrebe učenja sintetizatorja **RSDO**

Leksikoni z izgovorjavami

a) knjižno **RSDO**

b) narečno

Korpusi besedila za jezikovno modeliranje

Govorni viri so domensko odvisni:

a) splošni **RSDO**

b) medicina

c) pravosodje

d) ...

Aplikacijske domene:

a) pametni dom **RSDO**

b) razpoznavanje obrazov **RSDO**

c) drugo

Stanje pred in po RSDO



	PRED	PO
Govorna baza Cilj RSDO: razvoj javo dostopne govorne baze	<input type="checkbox"/> Ni javno dostopne govorne baze	<input type="checkbox"/> Na voljo javno dostopna govorna baza s 1000 urami govornih posnetkov
Podporna orodja Cilj RSDO: razvoj ključnih podpornih orodij	Ni javno dostopnih podpornih orodij za: <input type="checkbox"/> pripravo učnih podatkov <input type="checkbox"/> obdelavo rezultatov razpoznavalnika	Na voljo štiri podporna orodja: <input type="checkbox"/> sintaktični normalizator <input type="checkbox"/> akustični normalizator <input type="checkbox"/> fonemizator <input type="checkbox"/> punktuator
Razpoznavalniki Cilj RSDO: demonstracija uporabnosti podpornih orodij in govorne baze	Ni javno dostopnih modelov razpoznave.	Na voljo so: <input type="checkbox"/> dva splošna modela razpoznave <input type="checkbox"/> dva domenska modela <input type="checkbox"/> DEMO Portal

Orodje za postavljanje ločil - punktuator



- Orodje in API servis za postavljanje ločil in velikih začetnic
- Primer:

Glasbeni rokopis, ročno delo **L**udwiga van **B**eethovna, bodo vrnil dedičem najbogatejše družine na **Č**eškoslovaškem pred drugo svetovno vojno.

- Strojno učenje z uporabo globokih nevronske mrež (DNN):
 - Model 1: arhitektura RDNN, vektorske vložitve besed - **Glove**
 - Model 2: arhitektura RDNN za skupno predikcijo ločil in velikih začetnic, vektorske vložitve **SloBERTa**

Puntuator - rezultati



	Ločilo	Natančnost	Priklic	F1
Model 1	pika	0,862	0,869	0,865
	vejica	0,905	0,872	0,888
	vprašaj	0,722	0,527	0,609
Model 2	pika	<u>0,920</u>	<u>0,948</u>	<u>0,934</u>
	vejica	<u>0,956</u>	<u>0,953</u>	<u>0,955</u>
	vprašaj	<u>0,843</u>	<u>0,831</u>	<u>0,837</u>
	velika zač.	<u>0,985</u>	<u>0,980</u>	<u>0,982</u>

Sintaktični normalizator



- Na voljo dvosmerna pretvorba:
 - iz nenormaliziranega besedila v normalizirano
 - iz normaliziranega v nenormalizirano
- Primer:

Na včerajšnji tekmi so Angleži povsem nadigrali Senegalce in zmagali s **3:0**.

nenormalizirano

Na včerajšnji tekmi so Angleži povsem nadigrali Senegalce in zmagali s **tri proti nič**.

normalizirano

Sintaktični normalizator

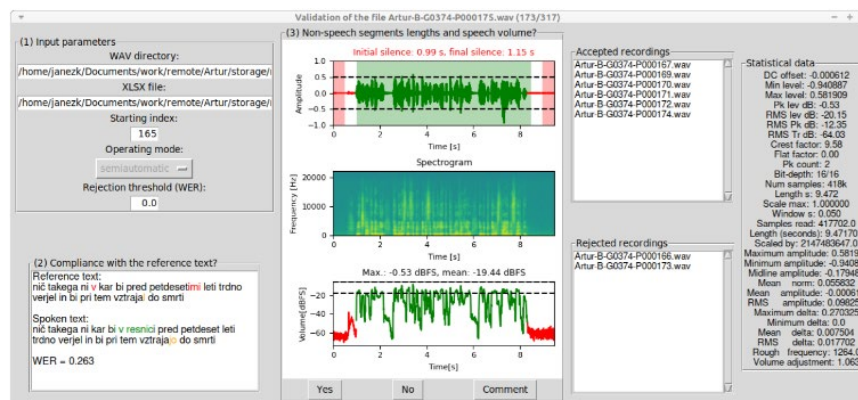


- Delovanje: stavčna analiza s pomočjo naučenih modelov NN (**Classla**) + pravila za posamezne kategorije.
- Podprte kategorije:
 - **navadna števila**: cela števila, vrstilni števniki, decimalne številke, ulomki, ure, razdelki, rezultati, telefonske številke, datumi, kompleksno trajanje, alfanumerične pojavnice, **rimskie števila**, **intervali**, **elektronski naslovi**, **simboli**, **enote**, **okrajšave**, **povezave** ...

Akustični normalizator



- Izdelane in preizkušene štiri modelne različice akustičnega normalizatorja z izvedbo metod strojnega učenja (**NRS**G, **SEGAN**, **DFL**, **NSNet**).
- Kakovost izboljšanih govornih posnetkov ovrednotena s tremi merami kakovosti (**PESQ**, **STOI**, **SISDR**), ki temeljijo na merjenju razmerja med signalom in šumom in tudi perceptivnemu vrednotenju kakovosti govornih signalov.
- Uporabljene mere kakovosti govornih signalov vključene tudi v orodje za ročno preverjanje posnetkov branega govora.



Fonemizator



- Izdelane tri modelne različice grafemsko-fonemskih pretvornikov (**Sequitur**, **Phonetisaurus** in **Deep Phonemizer**), ki temeljijo na izvedbi metod globokega strojnega učenja.
- S preureditvijo slovarja SloLeks 2.0 tvorjena učna in testna množica besed za globoko učenje izbranih modelov in ocenjevanje njihove zanesljivosti pri grafemsko-fonemski pretvorbi.
- Najboljši rezultat z modelom **Deep Phonemizer** in sicer 16,1% WER pri zahtevanem tvorjenju naglasnega mesta v besedah oziroma 8,6% WER brez zahtevanega tvorjenja naglasnega mesta.

Razvoj splošnih razpoznavalnikov



- Uporabljeni dva načina:
 - Z ločenim **akustičnim** in **jezikovnim modeliranjem**
 - Z uporabo **enovitega pristopa** (angl. e2e: end-to-end)
- Za jezikovno modeliranje uporabljena normalizirana **Gigafida**
- Evalvacija:
 - Razvita posebna testna množica, ki odraža strukturo Artur govorne baze
 - Testna množica izključena iz učne množice
 - Evalvacija izvedena na:
 - Testni množici Artur
 - Testni množici SloBench (za e2e)

Razvoj domenskih razpoznavalnikov



- Specializacija splošnega razpoznavalnika govora z **adaptacijo na ožjo domeno** (pametni dom, opisovanje obrazov).
- Namen: prikaz možnosti izboljšanja rezultatov razpoznavanja govora z uporabo ciljne domene.
- Viri za adaptacijo:
 - za domeno “pametni dom” podsklop govorne baze Artur, pripravljen.
 - za domeno “opisovanje obrazov” strojni prevod domenskega korpusa v angleškem jeziku.
- Evalvacija: ločeni testni seti

Kakovost razpoznavalnikov



- Referenčni razpoznavalnik: **WER 27%**
- Cilj RSDO: **WER < 15%** za splošni in **< 10%** za primera domenskih razpoznavalnikov
- Doseženi rezultati:
 - Splošni razpoznavalnik (ločena akustična in jezikovna faza): **WER: 11,66%**
 - Splošni razpoznavalnik z arhitekturo e2e: **WER: 5,58%** (SloBench: **WER: 6,79%**)
 - Domenski razpoznavalnik za pametni dom: **WER: 6,96%**
 - Domenski razpoznavalnik za opisovanje obrazov: **WER: 8,28%**

Dostopnost rezultatov <https://storitve.cjvt.si/>



- Splošni razpoznavalnik (Kaldi):
 - model: CLARIN
 - Servis API:
- Splošni razpoznavalnik (e2e):
 - model: [CLARIN](#)
 - Servis API: [CLARIN](#), [GitHub](#)
- Domenski razpoznavalnik 1:
 - model: CLARIN
- Domenski razpoznavalnik 2:
 - model CLARIN
- Puntuator:
 - model: [CLARIN](#)
 - Servis API: [CLARIN](#), [GitHub](#)
- Sintaktični normalizator:
 - orodje: [CLARIN](#), [GitHub](#)
- Sintaktični denormalizator:
 - orodje: [CLARIN](#), [GitHub](#)
- Akustični normalizator:
 - orodje: [CLARIN](#), [GitHub](#)
- Fonemizator:
 - orodje: [GitHub](#)

Ključni partnerji

- UL FRI, Laboratorij za podatkovne tehnologije
- UL FE, Laboratorij za umetno zaznavanje, sisteme in kibernetiko
- UM FERI, Laboratorij za digitalno procesiranje signalov
- VITASIS d.o.o.



VITASIS



Hitrost razpoznavne in poraba virov



Podatki za e2e model

	Hitrost	Poraba računskih virov
CPU	Faktor pospešitve: 1 do 1,5	servis + model: 2GB RAM izvajanje: 16 GB RAM za 300 s posnetek
GPU	Faktor pospešitve: 20 do 25 (Nvidia T4, 16 GB)	servis + model: 2GB RAM izvajanje: 3 GB RAM za 300 s posnetek

Demo

- Razpoznava s splošnim razpoznavalnikom 
- Avtomatsko postavljanje ločil

- Sinteza  

